Adult2child: Motion Style Transfer using CycleGANs

Yuzhu Dong\textsuperscript{1}, Andreas Aristidou\textsuperscript{2}, Ariel Shamir\textsuperscript{3}, Moshe Mahler\textsuperscript{4}, Eakta Jain\textsuperscript{1}

University of Florida\textsuperscript{1}, University of Cyprus\textsuperscript{2}, RISE Research Centre\textsuperscript{2}, The Interdisciplinary Center Herzliya\textsuperscript{3}, Carnegie Mellon University\textsuperscript{4}
Motivation

- Child characters have gained popularity in the animation and gaming industries.
- Most of the work focused on creating adult motions, but very few for child motions.

Russell from movie “Up”  Bonnie from “Toy Story 4″
Challenges

- Difficulties of motion capturing children
  - Lack of patience
  - Easily distracted
  - Hard to follow instructions
Challenges

• Given the difficulties of motion capturing children, can we just use adult mocap data on child characters?

• Can we convince the viewers that the motions are from children?

• Jain et al[2016] found that viewers can differentiate child motion from adult motion by viewing point light display videos.
Key Ideas

• Adapt adult motions to child motions that captures both the postures and the timing of child motions.

• Achieve this goal without temporally aligned data given that adult motions and child motions can be drastically different.
Contributions

• **Architecture:**
  • First to adapt cycleGAN for motion style transfer that can alter timing.
  • Redesigned the generators and the discriminators. *No temporal alignment.*
  • Additional loss terms to output natural and smooth motions.

• **Representation:**
  • Espouse joint angles as an animation-centric representation. *Facilitate character binding and skinning.*
  • Motion words to *encode temporal/spatial information.*

• **Dataset:**
  Released a high-quality optical mocap dataset of children.
Related work

Zhu et al.[2017][1]

Image style transfer with unpaired training data

Holden et al.[2016][2]

Learn motion manifold from a large dataset(six millions frames)

Aberman et al.[2020][3]

Image style transfer with unpaired training data

---


Our Novel Child Dataset: Kinder-Gator 2.0

- 8 children (5-10 years old). 9 adults: (18 years old and above).
- "Jumping jacks", "Throw a ball", "Walk", "Walk as fast as you can", "Hop scotch", "Punch", "Kick", "Jog", "Run as fast as you can", etc.
- 2-3 repetitions for each action type.
Examples from Kinder-Gator 2.0
Pipeline

1. Down-sample (120 -> 60fps)
2. Motion words (60 frames)
3. Convert to Quaternion
4. Normalize
5. Match training data by type
6. CycleGAN
7. Denormalize
8. Convert to Euler angles
9. Concatenate back to sequences
10. Apply smooth filters
11. Deploy
Overall Architecture

Adversarial loss
\[ \mathcal{L}_{G_{a2c}} = 0.5 * \mathbb{E}_{c \sim p(c)} [D_a(G_{c2a}(c)) - 1] \]
\[ \mathcal{L}_{G_{c2a}} = 0.5 * \mathbb{E}_{a \sim p(a)} [D_c(G_{a2c}(a)) - 1] \]

Cycle loss
\[ \mathcal{L}_{cycle,c} = G_{a2c}(G_{c2a}(c)) - c \]
\[ \mathcal{L}_{cycle,a} = G_{c2a}(G_{a2c}(a)) - a \]

Coherence loss
\[ \mathcal{L}_{coherence,a} = \sum_t \sum_{DOF} ||G_{a2c}(a)(t) - G_{a2c}(a)(t-1)|| \]
\[ \mathcal{L}_{coherence,c} = \sum_t \sum_{DOF} ||G_{c2a}(c)(t) - G_{c2a}(c)(t-1)|| \]

Transition loss
\[ y = G_{c2a}(c) \]
\[ \mathcal{L}_{transition,c} = \sum_t \sum_{DOF} ||y_t(t_{overlap:end}) - y_{t+1}(0 : t_{overlap})|| \]
Generator’s Architecture

Convolution convolves along the quaternion axis and the temporal axis
Discriminator’s Architecture

Patch GANs

<table>
<thead>
<tr>
<th>Layer</th>
<th>Filters</th>
<th>Kernel Size</th>
<th>Stride</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv1</td>
<td>64</td>
<td>4 x 4</td>
<td>2</td>
</tr>
<tr>
<td>Conv2</td>
<td>128</td>
<td>4 x 4</td>
<td>2</td>
</tr>
<tr>
<td>Conv3</td>
<td>256</td>
<td>4 x 4</td>
<td>2</td>
</tr>
<tr>
<td>Conv4</td>
<td>512</td>
<td>4 x 4</td>
<td>2</td>
</tr>
</tbody>
</table>

60 (frames) → 25 (joints) → 4 quaternions
Implementation Details

• We implemented and trained the model on Google Colab Pro with P100 or T4 graphics card.
• The model was written in Python using TensorFlow library.
• We trained the model for 180 epochs and the training takes ~7 hours.
• The trained model is 8.67MB.
Our Results: Punch

Input adult

Ours

Reference child
Our Results: Run as fast as you can

Input adult  Ours  Reference child
Compare with state-of-the-art: Walk as fast as you can
Compare with state-of-the-art: Jump as high as you can

- Input adult
- Holden et al[2016]
- Aberman et al[2020]
- Ours
- Reference child
Ablation Studies

No cycle loss        No coherence loss        With transition loss     Overall loss
adversarial+cycle+coherence
Perceptual Study

• Point light display of five conditions: child, adult, Aberman, Holden, ours
• 41 participants
• Does this motion belong to a Child or an Adult? (Child/Adult)
• Indicate the naturalness of the motion on a 7-point Likert scale.
Conclusion

• We have presented a method that allows adult2child style transfer,
• We have introduced two additional losses to condition the network, temporal coherency loss, and transition loss.
• The use of motion words helps the network to learn both the spatial and temporal information about motions.
Future Work

- Remove foot sliding/skating by adding constraints (inverse kinematics).
- Investigate mechanisms to change the sequence length and allow smooth blending.
- CycleGAN introduce noise in the output (unexpected angle). Explore additional constraints to reduce jitters.

Code download (in preparation):
https://gitlab.com/jainlab/cyclegan-1-master

Dataset download:
https://jainlab.cise.ufl.edu/publications.html#Adult2ChildCycleGAN

Yuzhu Dong (looking for an industry research position!)
yuzhudong@gmail.com